DOCUMENT RESUME

ED 372 366 CS 011 786

| | |
|---|---|
| AUTHOR | Price, Janet; Schwabacher, Sara |
| TITLE | The Multiple Forms of Evidence Study: Assessing Reading through Student Work Samples, Teacher Observations, and Tests. NCREST Reprint Series. |
| INSTITUTION | Columbia Univ., New York, NY. Teachers Coll. National Center for Restructuring Education, Schools and Teaching. |
| SPONS AGENCY | New York City Public Schools, Brooklyn, N.Y. |
| PUB DATE | May 93 |
| NOTE | 66p. |
| AVAILABLE FROM | NCREST, Box 110, Teachers College, Columbia University, New York, NY 10027 ($5 prepaid). |
| PUB TYPE | Reports - Research/Technical (143) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Classroom Observation Techniques; Elementary Education; *Evaluation Methods; Pilot Projects; Public Schools; Readability; *Reading Achievement; Reading Research; *Standardized Tests; *Student Evaluation; *Test Use |
| IDENTIFIERS | *Alternative Assessment; Degrees of Reading Power Test; New York City Board of Education |

ABSTRACT

A pilot study systematically collected classroom assessment data in reading and compared that data to the students' test scores on the Degrees of Reading Power (DRP). A total of 61 teachers in 18 New York City public elementary schools gathered the data for 268 students in grades 2 through 6. Teachers identified the books their children could successfully read by listening to the children read aloud, discussing the books, and evaluating student writing about the books. Teachers then looked up the DRP readability values of those books in published guides and compared them to the children's test scores reported in DRP units. Results indicated that the children with the lowest scores could read books that were 10 to 20 DRP units higher on the average than their DRP test scores. Results also indicated that the classroom assessment techniques used by the participating teachers yielded far more information about their students than a standardized multiple choice test such as the DRP. Findings suggest that teachers should collect multiple forms of evidence about student progress in a rigorous, systematic way and that this evidence has consequences both in the classroom and beyond it. This kind of assessment can inform and drive instruction in a positive way while at the same time make schools more accountable. (Appendixes present a letter of commitment, eight tables or charts of data, a sample record of reading performance, a student writing sample, and a sample miscue analysis form.) (RS)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 372 366

CS

NCREST Reprint Series

# *The Multiple Forms of Evidence Study*

## *Assessing Reading Through Student Work Samples, Teacher Observations, and Tests*

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D. Herrington
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

**Janet Price**
**Sara Schwabacher**
*The Fund for New York City Public Education*

with the assistance of
**Ted Chittenden**
*Educational Testing Service*

CS011786

National Center for Restructuring Education, Schools, and Teaching
Teachers College, Columbia University



BEST COPY AVAILABLE

The Multiple Forms of Evidence Study is sponsored by
**The Accountability Project**
A partnership of
The Fund for New York City Public Education
and the New York City Public Schools

The **Fund for New York City Public Education** mobilizes private-sector support for systemwide improvements in the city's public schools. The Fund's programs have supported school restructuring, professional development, and school/community partnerships in more than 500 public schools.

For more information on the Multiple Forms of Evidence Study please contact:

Maureen Houtrides
OREA
New York City Public Schools
110 Livingston Street
Brooklyn, NY 11201
(718) 935-3284

Sara Schwabacher
Fund for New York City Public Education
96 Morton Street
New York, NY 10014
(212) 645-5110

This report is reprinted and distributed by NCREST. Additional copies of this publication may be ordered for $5 each. All orders must be prepaid by check or money order payable to NCREST. Contact:

NCREST
Box 110, Teachers College
Columbia University
New York, NY 10027
(212) 678-3432
Fax: (212) 678-4170

The **National Center for Restructuring Education, Schools, and Teaching** (NCREST) was created to document, support, connect, and make lasting the many restructuring efforts going on throughout the nation. NCREST's work builds concrete, detailed knowledge about the intense and difficult efforts undertaken in restructuring schools. This knowledge is used to help others in their attempts at change, to begin to build future education programs for school practitioners, and to promote the policy changes that will nurture and encourage needed structural reforms. The Center brings together many voices: those of practitioners and researchers, parents and students, policy makers and teacher educators.

NCREST is supported by a major grant from the DeWitt Wallace-Reader's Digest Fund. NCREST's work in New York City, through its Center for School Reform, is supported by the Leon Lowenstein Foundation and the Aaron Diamond Foundation. Other funders have included the Center for Collaborative Education, the Danforth Foundation, the Fund for New York City Public Education, Impact II, the Lilly Endowment, Inc., the Metropolitan Life Foundation, the National Center for Research on Vocational Education, the New York Community Trust, the New York State Department of Education, and the Regional Laboratory for Educational Improvement of the Northeast and Islands.

# *The Multiple Forms of Evidence Study*

## *Assessing Reading Through Student Work Samples, Teacher Observations, and Tests*

**Janet Price**
**Sara Schwabacher**
*The Fund for New York City Public Education*

with the assistance of
**Ted Chittenden**
*Educational Testing Service*

May 1993

# Acknowledgements

The Fund for New York City Public Education gratefully acknowledges the many people who made the Multiple Forms of Evidence Study possible, including:

**Our colleagues who collaborated in organizing the pilot study:**

Dr. Jill Benado, Suzanne Marten, and Jane Spielman
Center for Educational Options, City College

Dr. Ted Chittenden and Dr. Jacqueline Jones
Educational Testing Service

Dr. John Schoener and Maureen Houtrides
Office of Research, Evaluation, and Assessment
New York City Public Schools

**The teacher/consultants who reviewed and summarized the data:**

PS 84M
Marilyn Bell, Assistant Principal
Mary Foote
Diane Glenn
Clarence C. Loftin, III
Linda Margolin

PS 261K
Mildred DiStefano
Carmen Fulford

**And, above all, the 61 participating teachers, their 18 principals, and the SBM/SDM facilitators, OREA assistors, and District Office staff who assisted their efforts.**

The Multiple Forms of Evidence Study was made possible by generous grants from the Bruner Foundation, the Robert Sterling Clark Foundation, the Leon Lowenstein Foundation, J.P. Morgan, and others.

## Executive Summary

This pilot reading assessment study is sponsored by the Accountability Project, a partnership between the Fund for New York City Public Education and the New York City Public Schools' Office of Research, Evaluation and Assessment. The goal of the Accountability Project is to strengthen student assessment and school evaluation practices citywide.

In the spring of 1992, 61 teachers in 18 elementary schools systematically collected classroom assessment data in reading for 268 students and compared that data to the students' test scores on the standardized, multiple choice reading test administered that spring, the Degrees of Reading Power (DRP).

The teachers took advantage of a useful feature of this test, which reports student test results in "DRP units" designed to correspond to the relative difficulty of reading matter. Teachers were able to identify actual books their children could successfully read by listening to the children read aloud, discussing the books with the children and evaluating student writing about the books. Teachers then looked up the DRP readability values of those books in published guides and compared them to the children's test scores reported in DRP units.

The resulting findings are dramatic. The children with the lowest scores could read books that were 10 to 20 DRP units higher on the average than their DRP test scores. This meant that the actual reading performance of many of these children was at least at grade level, although their test scores were well below. When data were aggregated, significant differences were found between the average test results and the average difficulty of books the group could read successfully. For instance, although only 38% of the third graders in our sample scored at or above grade level on the annual test, fully 60% of these same students could read books appropriate to their grade level. In part, these results are an artifact of standardized measurement. For example, it is recognized that in general, standardized tests are less reliable for the highest and lowest scoring students.

Whatever the reasons for the discrepancies, the important lesson here for practitioners and policy makers is that it would be inaccurate and potentially harmful to judge either individual students or schools on the sole basis of a multiple choice test given once per year. Unfortunately, because standardized tests are the only mandated assessment measure, they often are used to track students by achievement level and to make other "high stakes" decisions about individual students, even when students' classroom performance is better than the test results would predict. Inappropriately low expectations and lack of challenging work can result in such students never achieving their full potential.

Moreover, aggregate test scores are commonly used not as one factor, but as the sole factor for judging school performance. Indeed, the school system is required by state decentralization law to rank schools annually on the sole basis of their reading scores. Many high stakes decisions about schools are made based on test scores. For instance, schools with

low DRP scores in the third, sixth and/or eighth grades wind up on the State Education Department's list of "Schools Under Review." Such over-reliance on test scores hurts accountability because, as this study illustrates a single test score does not yield an accurate or complete picture of achievement even in the subject tested.

A second important finding of this study is that the classroom assessment techniques used by the participating teachers yielded far more information about their students than a standardized multiple choice test given one day in the spring. These assessment practices gave teachers rich and immediately useful information about the reading strategies their students used and which literacy skills should be the focus of further instruction. The teachers in this study found it well worth their while to collect and compare different types of evidence of their students' progress because it helped them serve their students better. In particular, they found that these methods were essential to ensure that students did not "fall through the cracks." As one teacher noted, "With traditional tests, you teach the lesson, not the student. But with this work, I get to know the students and whether I am getting through to them." Most indicated that they planned to continue this work in the 1992-93 school year, starting earlier and working with more children. Given this feedback, the Accountability Project is expanding this study to additional schools.

In the interest of more accurate assessment and better instruction, it is urgently important that teachers collect multiple forms of evidence about student progress in a rigorous, systematic way and that this evidence have consequences both in the classroom and beyond it. Specifically, collection and use of multiple forms of assessment evidence should have three main consequences:

1. Informing instructional strategies within the classroom;

2. Traveling with students to inform instructional and placement decisions about students as they move to a new grade or new school;

3. Serving as an indicator of overall school effectiveness.

There should be a strong message at every level of governance that responsible use of multiple forms of assessment is a necessary characteristic of any good school and over-reliance on any one piece of evidence is bad practice.

This study has important implications for the city, the state and the nation. As President-elect Clinton focuses on national education goals and the New York State Regents consider proposals for curriculum and assessment reform, it is critically important for policy makers to understand that standardized tests alone do not yield a reliable picture of either student achievement or school effectiveness. If our goal is to produce students who can read with enjoyment and understanding and who can talk and write thoughtfully and clearly, assessments must examine students' abilities to do those things.

A number of promising assessment pilots are underway in New York City. Over 50 schools are piloting the Primary Language Record, a comprehensive assessment system for reading, writing and oral communication for elementary schools. The school system has developed an early childhood reading assessment and a middle school performance-based mathematics assessment for use citywide and plans to expand such strategies to additional grades. Both the city and New York State are participating in the New Standards Project, a national effort to develop performance-based assessments in mathematics and communication arts. Yet despite these major steps forward, standardized multiple choice testing still dominates decision making at every level.

What 61 teachers did in a pilot study could and should ultimately be standard practice. In some schools it already is. If multiple forms of evidence were given the importance that test scores alone are afforded today, teachers would be willing and able to collect and use that evidence. This is the kind of assessment that can truly inform instruction and drive instruction in a positive way while at the same time making schools more accountable.

## Introduction

In the Spring of 1992, the Multiple Forms of Evidence Study engaged 61 teachers in 18 schools in the systematic assessment of their students as readers. These teachers collected evidence on a total of 268 students in grades two through six. Most teachers concentrated on studying three to five students, but five teachers collected evidence on their entire classes.

This work was part of the Accountability Project, a partnership between the Fund for New York City Public Education and the New York City Public Schools' Office of Research, Evaluation and Assessment (OREA). The goal of this partnership is to bring the best student assessment and school accountability practices to the New York City public schools.

Both the Multiple Forms of Evidence Study and the Accountability Project as a whole are premised on the belief that responsible and accurate assessment must be broadly based to include evidence of student progress, collected over time and across classroom settings. Sources of evidence include:

- Teacher observations
- Periodic samples of student work and performance
- Standardized and classroom test results

These first two forms of evidence are central to teachers' planning and day-to-day instruction. They offer a fuller and more accurate picture of student progress than a single test taken on a single day. Such evidence looms large in experienced teachers' evaluation of student reading. Yet rarely are these classroom-based forms of evidence taken into account in important policy and placement decisions about individual student placement, staff assignments, classroom configurations, resource allocation, or the effectiveness of schools and districts. Too often, these important decisions are made on the basis of a single test score, usually on a multiple choice standardized test.

The Multiple Forms of Evidence Study sought answers to three key questions:

- What are the most useful ways for teachers to document student learning?
- How can such classroom evidence inform assessment decisions beyond the classroom?
- In what ways can such evidence be compared to test results?

The study had two main objectives:

- To involve teachers in the systematic collection and review of classroom-based evidence, as a first step towards broadening the base of school and district assessment practices.

- To compare classroom assessment results with standardized test scores using the readability values of books the children were reading as a "bridge" between teachers' records and test scores.

The Degrees of Reading Power (DRP) is the standardized test given every spring to New York City public school students beginning in the second grade. It is also the test the New York State Education Department requires for all third, sixth and eighth graders statewide. The DRP is used for many high stakes decisions. The state uses test results, along with a few other indicators, as the basis for publicly identifying "Schools Under Review"--that is, poor performing schools. State decentralization law requires that New York City schools be ranked according to their performance on the annual reading test. Test results are often used to evaluate the effectiveness of schools and districts and their principals and superintendents. The reading test is also commonly used to place students in classes for high or low achievers.

DRP test results, like most standardized tests, are generally reported to the public in percentiles. If Johnny scores at the 50th percentile, it means half the students taking the test did worse than Johnny and the other half did better. Students who score at or above the mean are frequently described in public reports as reading "at or above grade level." But in fact, this way of reporting scores merely compares students to each other and does not give any information about what the student can or cannot read.

However, DRP results are also reported in "Degrees of Reading Power" units which are intended to correspond to the "Degrees of Reading Power" readability values of books. Readability values for many books commonly found in public school classrooms are available in a guide prepared by the publishers of the DRP test. Thus, one way DRP test scores can be used is to identify appropriate reading material for individual students.

The teachers in the Multiple Form[illegible] Evidence Project did the reverse. They first used classroom assessment methods to identify books the students could successfully read and only then looked up the books' DRP values. The readability values of these books were then compared to the child's DRP score on the spring 1992 test. The existence of the readability value guides made it possible to compare alternative assessment evidence to test scores.

A small group of teachers first used this approach to assess their students' reading during the 1990-91 school year. With the assistance of Dr. Ted Chittenden, a research psychologist with the Educational Testing Service, the Accountability Project expanded the "Multiple Forms of Evidence" work as a sensible and cost effective way to involve additional schools and teachers in the exploration of assessment techniques.

During the second semester of the 1991-92 school year, the Fund and the school system recruited schools to participate in a pilot Multiple Forms of Evidence Study. We decided to focus on grades three through six, for several reasons. Teachers in those grades had, by and large, not been reached by other assessment pilots, yet those were excellent grades for using many types of classroom assessment in reading. These are also grades in which test scores often currently play a central role in placement decisions, particularly when students make the transition to a middle level school.

Chapter One of this report describes the organization of the study and the nature of the data collected. Chapter Two focuses on the patterns that emerged when test scores were compared to the readability values of the books children actually read. Chapter Three describes the types of assessment evidence that teachers found most useful and how that evidence proved useful, both in assessment and in instruction. The final chapter summarizes the policy implications of this pilot study.

# Chapter One

## Methodology

This chapter describes the organization of the Multiple Forms of Evidence study and the nature of the evidence collected by teachers. This background is intended both to help readers understand how the findings set out in subsequent chapters were obtained and to provide "how to" information for educators and policy makers who would like to adopt our methodology.

### Organization of the Study

In the Fall of 1991, a team was convened to organize the study that included staff from the Fund and OREA, researchers from Educational Testing Service, staff developers from the Center for Educational Options and teachers who had participated in the early version of the study. The team decided to concentrate on grades three through six and to invite any New York City public school serving any of those grades to participate. Participation was conditional upon a minimum of three teachers in that school being willing to make the following commitments:

1. To collect evidence on at least three students, using classroom assessment techniques to identify approximately three books each child could read with comprehension;

2. To meet regularly with other participating teachers within their school during the spring semester to reflect on what they were learning about their students through classroom assessment and to hone their assessment techniques.

Interested schools and districts were invited to an introductory session in March of 1992. They were briefed on the types of evidence that constitute "multiple forms" and heard from teachers who had participated in the earlier study. Participating schools were asked to submit a letter of commitment (see Appendix A) and teachers were asked to use the "Record of Reading Performance" form specifically developed for this study. A blank copy of the Record of Reading Performance form is included at the end of this report as Appendix N, which may be duplicated for classroom use.

Three groups provided invaluable help in recruiting schools and in convening and facilitating staff development meetings of participants within schools. The first were OREA staff and consultants already deployed through the Accountability Project to work on site with schools exploring alternative assessment strategies. The second were the School-Based

Management/Shared Decision-Making facilitators already deployed to assist schools involved in that initiative. The third were school district staff, particularly assessment liaisons who were already in regular contact with OREA. Special staff development sessions were held for these groups, to prepare them to better assist teachers involved in the study.

Principals and assistant principals also played a critical role in participating schools, providing a wide range of administrative support. For instance, in many cases, they rearranged schedules to provide participating teachers with common prep periods to facilitate their meetings.

Two groups of teachers participated in the study. The first group, termed "Small Group Studies," included many teachers who were formally collecting "alternative assessment" data for the first time, through this study. They documented the reading activities of a small sample of children in their classrooms, usually three children. They were encouraged to choose children who represented the range and variety of readers in their class. Teachers often included in the group children who read "fairly well" -- in the teacher's judgment -- but who tend to score poorly on tests.

The other group, "Whole Class Studies," included five teachers who undertook documentation of reading performance for all children in their classrooms. Since all five classes were heterogeneous in nature, the teachers' findings give a picture of the full range of student achievement. The five teachers' experience was also invaluable for the insights they provided on how a teacher can organize time and resources to formally collect classroom assessment data for an entire class. [See Appendix B for a breakdown by grade level of the participating classrooms.]

## Data Collection

Teachers participating in the "Small Group Studies" recorded data on a form prepared for the project, Record of Reading Performance. The following categories of information are represented in the form:

[a] the titles of books the child was actually reading around the time of testing [column 1]

[b] teacher's judgment about the child's abilities to read and understand the books [columns 2 and 3]

[c] the DRP Readability Values of the books, as listed in DRP publications [column 4]

[d] an indication of the nature of evidence used by the teacher to evaluate the child's reading [columns 5-9]

[e] comments by the teacher concerning the child's reading [column 10]

[f] attachments: teachers were asked to attach student work samples or teacher notes as examples of their methods for documenting reading. These might include writing samples, notes taken while listening to a student read out loud, or interview notes.

[g] scores from the DRP test administered in Spring 1992.

A total of 54 teachers representing 16 schools collected information of this kind for 150 students. See Appendix C for a completed Record of Reading Performance form including the teacher's attachments.

The teachers participating in Whole Class Studies compiled reading records for each student in their classroom consisting of:

[a] titles of books the child was reading around the time of testing

[b] DRP readability values for those books, as listed in DRP publications

[c] DRP test scores, Spring 1992

Five teachers, from two different schools, participated in this aspect of the project, collecting information for a total of 119 students. The teachers all taught multi-age classrooms, which combined two grade levels. One teacher taught a 3rd/4th grade classroom and four teachers taught 5th/6th grade combination classrooms.

At each site, a facilitator met with teachers on an average of three times during the project. These meetings were occasions for teachers to share their work, discuss issues, and plan next steps. The study was designed as a "decentralized" project in the sense that the most important discussions would occur at the local building levels, in meetings of teachers, facilitator and others. Topics for these lively and collegial meetings included evaluation of the merits of the various assessment strategies, discussions of how to use the resulting information in instruction and brainstorming on how to overcome the various barriers teachers encountered. Facilitators' notes from these meetings were an invaluable source of information for this study. Some of the data from these meetings are incorporated in Chapters 3 and 4 of this report.

## Data Analysis

At the end of the school year, 61 teachers from 18 schools turned in their data on students and their written notes and observations on their assessment work. To protect the confidentiality of student records, surnames and other personal identification of children were not included. Facilitators and OREA consultants also turned in their notes from school meetings. Over the summer, these data were reviewed and summarized by a team of teacher consultants assisted by the study organizers. The information was summarized along two lines: a comparison of test scores and readability indices [See Chapter 2] and analysis of teachers' assessment practices [See Chapter 3.]

All participating teachers were invited to a follow-up meeting in November, 1992 and were asked to review and comment on a draft of this report. At that time participants also discussed how they were continuing the work during the 1992-93 school year, the obstacles they faced and their recommendations of strategies for addressing these obstacles. Many of their comments have been incorporated into this document.

## Chapter Two

## Comparing Indices of Reading Performance: Test Scores and Readability Values

The Degrees of Reading Power tests (DRP) are a series of standardized reading comprehension tests used by the New York City school system for grades 2 and up. The tests consist of a sequence of passages in increasing order of difficulty. The test's scores, expressed in DRP units, represent an estimate of the most difficult prose that a student can read with understanding; the estimate is based on the student's multiple choice responses to the passages.

In addition to test scores, the DRP also publishes readability values for instructional materials and children's books that may be found in classrooms. The readability values are derived from a formula which takes into account the surface features of prose. Thus prose that is characterized by simple, common words and short sentences is generally easier, or more "readable," than prose with complicated structures and uncommon words. While the scale ranges from 15 to 100 points, most prose actually ranges from 30 units, which is the sort of text found in first-grade basal readers, to 85 units, the sort of prose encountered in graduate school. Appendix D, an excerpt from a DRP publication, provides some illustrative samples.

One of the unique features of the DRP is that it uses a common scale to estimate both the reading ability of students and the readability of prose. Thus the same scale -- expressed in DRP units -- is used to score a child's test performance and to estimate the difficulty or readability of material that the child might read. According to the DRP guide [1989,] one benefit of the common scale is that it allows a direct comparison of test scores and readability estimates. The guide suggests that such a feature enables teachers to select books that are "matched" to a student's reading abilities.

## DRP Data Used in Pilot Study

The portion of the pilot study described in this chapter explored ways in which the DRP scoring system might be used to compare classroom-based evidence with test-based evidence concerning children's reading achievements. Essentially, using DRP readability values as a bridge, comparisons were made between classroom evidence of reading activities [i.e. what books the child was actually reading] and evidence or estimates derived from test performance.

As described in Chapter 1, the teachers' documentation of children's reading activities included compiling a record of books the child was reading during the Spring, around the time of testing. The readability values for these books, expressed in DRP units, served as one

indication of the level of the child's reading accomplishments. These values could in turn be directly compared to the child's scores on the DRP test.

## Test outcomes

The DRP Mid-instructional score was selected as the key indicator from the test reports. Students receive several different DRP unit scores. The lowest score, "the independent level," suggests the difficulty level of books that students can read with comfort on their own. The "mid-instructional score" suggests the level of difficulty students can manage when reading classroom materials with some instructional support from the teacher. The highest score "frustrational level," suggests the level of difficulty students might find very challenging. The Mid-instructional score was considered the most appropriate for present purposes since it would seem to indicate the level of material that is appropriate for most instructional purposes. That is, material at the mid-instructional level is presumably within the child's understanding and grasp, although the child may still need some help from the teacher.

## Book readability values

The teachers had recorded the titles of books the child was reading on the form provided by the project. Readability values for these books were then taken from the DRP publications that provide such data for more than 1,500 children's books. In the majority of cases, more than one title was listed in a child's record. For purposes of data analysis, the Median Readability value was used as an indicator of the readability of the materials in that child's record.

## Sample size

The following analysis is based on data in the records for 207 children; 88 from the Small Group Studies and 119 from the Whole Class studies. When categorized by grade level, the sample includes 78 middle-grade [3rd/4th] and 129 upper-grade [5th/6th] students. (The total sample is somewhat smaller than the overall sample described in Chapter 1 because some records lacked readability values and could, therefore, not be included in this particular analysis. This occurred where the particular books the student read did not appear in DRP Readability Guides.)

## Note concerning characteristics of the sample

Before discussing the findings, characteristics of the samples of students need to be noted. This was a pilot study dependent upon voluntary participation of teachers. Moreover, in the Small Group Studies, no attempt was made to control the selection of students other than to ask

teachers to select students who represent the range of reading activities in their classrooms, including those who might perform poorly on conventional measures.

Appendix E shows the distribution of DRP test scores [norm-referenced scores] for students included in the present analysis. Scores are grouped according to quartile. A slight majority of students obtained scores below the 50th percentile [i.e., the 1st and 2nd quartiles.] As commonly translated in the newspapers, these would be students "reading below grade level" because their scores fall below the "cut-off" point of the 50th percentile.

For the Small-Group Study, the even distribution across the four quartiles suggests that this sample of students did indeed represent the full range of tested reading proficiency. In the case of the Whole-Class study, the full range is also represented although with a proportionately greater number scoring in the lower quartiles.

Given the similarity of the distribution of results between the small group and whole class study samples, the two groups are combined in subsequent analyses of results. The data obtained from three teacher-selected students proved to be similar and comparable to the data collected from teachers who studied all the students in their class.

## Results

### Data trends

The overall trends in the data are depicted in Appendix F, which shows average test score and readability values for each grade level. It should be emphasized that the test scores [mid-instructional] in this Figure are expressed in DRP units, which are criterion-referenced. Therefore, as would be expected, there is a steady rise in test score averages across the grades since students read progressively more difficult material as they get older. A criterion-referenced score, such as the DRP Mid-Instructional score, reflects such change since it compares student test performance to an external criterion, text difficulty. [A norm-referenced score, by contrast, compares students to other students within the same age group, and would not directly reflect such growth. In fact, for the current sample, the average norm-referenced score consistently places the groups for each grade level slightly below the 50th percentile.]

In contrast to the rise in test scores, the trend across grade levels for average readability values is much "flatter." In aggregate, the books that the children were reading at 3rd grade appear comparable in surface difficulty to the reading diet of 4th graders. Similarly, the average readability values for 5th and 6th grades are much the same. And indeed the overall difference between the middle [3rd/4th] and upper [5th/6th] grades is slight, on the order of 3 or 4 DRP units. It should be noted, however, that while such an increment is slight it does in fact represent a discernable rise in text difficulty, as illustrated in Appendix D. It might also be noted that the average readability values for each grade in Appendix F are very close [within 2 or 3 units] to averages from a state-wide sample of instructional materials for each grade, as reported in the DRP guide [p. 10].

**Individual data: Difference between readability values of books the students could actually read and student test scores**

One of the reasons prompting teachers' initial interest in participating in the Multiple Forms of Evidence study was their concern about discrepancies between their own judgments of a child's reading achievements and a test's estimate. This sort of concern is often expressed regarding children who are quite competent in their everyday reading in the classroom, yet perform poorly, relatively speaking, on tests. Sometimes this is attributed to poor test-taking skills or to the child's lack of interest in figuring out DRP passages. The teachers also note that the DRP passages place emphasis upon nonfiction, while much of children's every day reading experiences are with fiction. Whatever the reasons for discrepancies, concerns of this sort are voiced every year by teachers.

As one way of examining this question, a "difference score" was derived from each student's record by calculating the difference between the Median Readability value of books the child was reading and the child's test score [Difference score = Readability - Test [Mid]]. Appendix G shows the percentage distribution of difference scores, for middle grade and upper grade children in the sample. For convenience, the scores are grouped in intervals of 10 DRP units.

Appendix G indicates that for about half of the students, the Readability and Test estimates are approximately the same; that is, they fall within a range of plus or minus 10 DRP units. This accounts for about 45% of the 3-4 graders and 55% of the 5-6 graders. For this group, teachers' findings are fairly consistent with the students' test scores. It should be noted, however, that a disparity of 10 DRP units already represents a discernible difference in the difficulty of a text.

But this table also indicates that for about a quarter of the students, the differences between scores are in the positive direction and greater than 10 units. [Most of the difference scores in this group fall within the range of 11 to 20 units, but some even exceed 20 units.] These are students who are reading books that are significantly more difficult than is predicted by test results. The table indicates that this sort of discrepancy is more pronounced in the middle grades than in the upper grades.

Finally, the table represented in Appendix G indicates that there is another group of students, approximately 20% of the total, with Difference scores in the negative direction, below-10. These are students who are reading books that appear to be easier than predicted by their test scores. This sort of discrepancy is more common in the upper grades.

**Difference scores and percentile rankings**

Another way to look at Difference scores is to group the students according to their norm-referenced ranking. Appendix H is a figure depicting the distribution of the Difference

scores for the entire sample, combining the grade levels. In this Figure each student's difference score is plotted according to the student's percentile standing on the DRP test.

When viewed in this manner, the data indicate a systematic but changing relationship between the Difference Score and percentile rank. For children who score relatively poorly on the test [in the lowest quartile, 1-25 np] the difference scores tend to be large and in the positive direction, suggesting that they are testing "below" their observed reading estimates, or, -- expressed another way -- they are reading material that is more difficult than would be expected, given their low test scores. By contrast, for many of the good test takers [scoring in the 4th quartile, 75-99 np] the difference scores suggests they test "above" their observed reading achievements. For the students in the middle of the test range, 2nd and 3rd quartiles, the difference scores are not as great although they are consistent with the pattern set by the extremes.

The difference scores are aggregated in Appendix I -- via averages -- to show trends by grade level and quartile. The pattern noted in Appendix H is consistent for each grade, with the differences between observed and tested reading achievement being greatest at either end of the test score range. But the Figure in Appendix I also indicates some distinctions among grades, with the direction of difference generally more positive for the third and fourth graders, more negative for the upper grades. The third and fourth graders who scored in the bottom quartile on the test could read books an average of 20 DRP units higher than their test scores. An increase of 20 points on the DRP test would bring a student from well below grade level to above grade level. The fifth and sixth grade students who scored in the bottom quartile on the test could read books an average of 10 DRP units higher than their test scores, a significant difference.

## Reporting Aggregate Data

The data collected in this study hold implications for assessment of schools as well as assessment of individual children. Test scores are used by state and district policy makers to judge schools and educational programs. Aggregate test scores are commonly reported to policy makers and the public by listing the percentage of students in a school or at a grade level with norm referenced scores at or above the 50th percentile. Appendix J compares the percentile ranking of the aggregated data for third graders from this study and the percentage of students who read books with a DRP readability score appropriate for their grade level. For this table "appropriate for their grade level" is defined as being within two DRP units of the state average for the DRP readability values of instructional materials used in third grade classes (48 DRP units) as reported in the DRP guide.

Although only 38% of the third graders in our sample scored at or above grade level on the DRP test, fully 60% of these same students could read books appropriate to their grade level. If this group of third graders were a school on the most recent annual rankings report, this difference would be enough to move it up from 451st place to 203rd place. The New York

State Education Department relies heavily on aggregate third grade performance on the DRP in compiling its list of schools under review for poor performance. If it also considered the actual reading performance, as opposed to test performance, for third graders, it might well draw different conclusions about many schools.

## Discussion

The data analyzed in this chapter are statistical indicators. As such, they are estimates or quantitative proxies for phenomena that are inherently complex, i.e. the process of reading and the nature of text difficulty. As with many such indicators, they can be informative if not taken literally. Thus it is important to remember that the DRP test score is an estimate of reading achievement, which is based upon the student's abilities to deal with passages [mainly expository] taken out of context. In the same vein, the DRP readability values reflect only the surface characteristics of text, i.e. the "technical" difficulty as distinct from the "conceptual" challenge of the material. This distinction in itself became a matter of interest to teachers as they looked up readability values and became bemused by the fact that -- for example -- Hemingway's Old Man and the Sea [50] had essentially the same score as Old Yeller [51.], no doubt reflecting Hemingway's terse style of writing.

Undoubtedly, some of the patterns noted in the results reflect constraints of the measures. [In fact, there is a thorough discussion of the limitations of readability formulae in the DRP publications.] For example, the actual observed range of Readability values was much narrower than the corresponding range of test scores, although "in theory" the ranges should be similar since they are based on the same scale. Given such discrepancy it is not surprising to find large differences between the two scores at the extremes of the distributions.

These differences may also reflect the realities of elementary instruction. Children in these grades, whatever their apparent skills as readers, bring the knowledge and interests of their age to their reading experiences. Thus, those who score in the lower quartile are not "beginning readers," in need of books with very primitive text structures [e.g. DRP readability in the 30's.] Instead, they are 8, 9, or 10 year-olds with the knowledge and interests of that age. Conversely, for children who hit the ceiling of the test, attaining DRP Mid-instructional scores in the 70's and 80's, it does not follow that they should be matched to text of comparable difficulty, somewhere at the college or graduate school level.

It is also quite possible that the role of prior knowledge looms larger at the extremes of the range of reading skills. That is, beginners or struggling readers depend heavily upon such knowledge in gaining access to print. They especially need books that respond to their background knowledge and experience, as it may be the meaningfulness of such material that supports their tentative skills to make sense of the print. And, at the other extreme, skilled readers choose material that responds to their questions and interests -- they don't select material that "matches" their abilities to deal with surface features of text.

## The reading diet

The group data trends depicted in Appendix F appear to highlight two quite different aspects of children's development as readers. On the one hand, the steady rise in test scores implies growth in children's competence as readers over the course of the four elementary grades. But on the other hand, the flatter slope of the readability averages suggests that the reading "diet," which supports this growth, does not change as dramatically, at least at the level of the surface features used to determine readability values.

Putting aside for the moment the fact that readability values reflect only the surface features of prose, the indication of a gradually changing reading diet across these grade levels would seem to be in line with teaching practices in many of these classrooms. First, a number of teachers noted that some of the more popular children's books [whatever their readability value] could be profitably read by many of the children in the classroom, whether they were strong or wobbly readers. A good children's book has appeal and literary value for a variety of readers. Moreover, certain books were "passed around" the classroom, reflecting the strong social nature of reading and the children's interests in sharing experiences. Many of the classrooms also featured a class library that offered a variety of books to meet a variety of interests, and at the same time reflected some range of difficulty, from "easy-to-read" to more difficult. And while "good" readers might well select more difficult material than their less skilled classmates, it was also the case that these same children would sometimes read "easy stuff" for a variety of reasons. Conversely, children who were struggling as readers would sometimes tackle difficult material, because of compelling content.

These several qualities of the classroom reading program mean that the "diet" for reading experiences was a varied diet, one that would indeed change gradually over the grade levels. Unlike some conventional basal reading programs, in which reading material is tightly sequenced and prescribed according to a presumed level of reading skill, the reading experiences in many of the participating classrooms were marked by choice and variety. Hence, in retrospect, the flatness of the readability trend is not so surprising.

## The "match" between reader and text

At a general level, the expectation that teachers can help children select appropriate reading matter is certainly appropriate. And it is certainly true that teachers take the surface features of text into account [grammatical complexity, word difficulty, layout etc.] when selecting books for classroom instruction. In fact, many teachers arrange displays of books in the classroom libraries along an "easy-to-read" continuum.

The DRP supports this important responsibility of teachers by offering an additional way to evaluate books, and thereby promoting a better "match" between the reader's skill and text difficulty. However, since the DRP readability values do not take into account the actual content

or subject of the text, teacher judgment and teacher understanding of the process of reading remain critical.

The past 25 years of research on reading have contributed to a major shift in our theoretical knowledge, in which reading is viewed as an interactive process -- the reader shaping the text as much as the text shaping the reader. What readers bring by way of background knowledge and interest is critical in defining the reading experience.

Young readers do, of course, have to deal with the surface features of print in more explicit ways than do older, more proficient readers. But research and classroom experience has shown that for the beginning reader as for the adult, the search for meaning is paramount. Even while the child struggles to figure out the words, it is the meaning of the task that sustains the effort. The challenge for teachers at the elementary grades is to support and encourage the search for meaning, while at the same time help children develop the needed strategies and capabilities to deal with print in all its forms. This means that children sometimes "learn to read" by re-reading and returning to something that is familiar and easy; at other times they learn to read by tackling something more challenging because of its compelling content. While the surface features of text are a consideration in "matching" a book to a reader, the vital connections reside elsewhere, such as the discovery of ways the book or story respond to the child's interests and ideas. It is at this juncture that teachers' observations and sensitivity to children become so important. To quote from the DRP guide, "As with the factor of interest, teacher judgment is the most effective tool for assessing students' background and readiness for a given text." (1989, p.29)

## Conclusion

This pilot study explored ways in which DRP information could be used to compare classroom-based evidence with test evidence. DRP readability values were used as a "bridge" between classroom records and test results. Readability values, as estimates of the difficulty of material the child was reading, were compared with test scores, as estimates of the child's reading ability. The data indicate that for half of the students in the sample there was a difference of more than 10 DRP units between their test scores and the average readability value of books they could read with comprehension. The gaps between readability values and test scores were greatest at the extremes. Of particular interest is the striking finding that many low-scoring students can successfully read much more difficult books than their test scores predict.

Because readability values reflect only the surface difficulty of texts, it would probably not be worth the effort to routinely look them up and include them in all students' reading records. As discussed in the next chapter, teacher efforts are best concentrated on collecting and using classroom assessment data. However, when a teacher believes that DRP test results significantly underestimate a student's reading achievement, DRP readability values can serve an important function. Determining the readability values of books that the child can actually

read provides a way of summarizing the classroom evidence in the language of test evidence to facilitate a comparison. This provides a more complete assessment picture which should prevent inappropriately low expectations for students based on a single test score.

# Chapter Three

## Forms of Evidence

This chapter is concerned with the two sources of evaluation data gathered in the classroom -- samples of student performance and reflections by teachers and facilitators on their observations. We will examine the specific forms of evidence teachers used to determine the books their students could read and what teachers learned from these forms of evidence.

In contrast to the multiple choice question and answer format used on standardized tests, teachers assessed the level of books their students could read by listening to them read aloud, discussing the books with individual children or in small groups and by evaluating students' writing about the books. Collecting this assessment evidence confirmed and supplemented teachers' informal observations and gave them rich, immediately useful information about the reading strategies their students used which led directly to practical ideas for further instruction.

The following sections describe 1) which forms of evidence teachers found most useful, 2) what teachers learned about individual student's reading skills, strategies, processes and attitudes and 3) the implications teachers found for their own instructional, pedagogical and evaluation practices.

### What assessment strategies were used by teachers

The fifty-four teachers who participated in the Small Group Study reported the strategies they used by completing a "Record of Reading Performance" for at least one student in their classroom (see appendix C). Appendix K reports the Nature of Evidence used to determine what books their students were able to read (eg. oral reading, group discussion, writing sample) recorded by teachers on this form. Appendix L shows the number of teachers who included attachments (eg. student work sample or formal miscue analysis as illustrated in Appendix M) with their record. A form of evidence is listed on these tables if it is self-reported by a teacher for at least half of the students he or she studied.

The great majority of teachers (92%) listened to a child read aloud. Eighty-seven percent also held an individual conference with the child. Although these strategies were widely used at all levels, they were used most at the middle grades and relied upon less in the upper grades. (96% of 3rd grade teachers and 100% of 4th grade teachers used oral reading as compared to 87% and 82% of 5th and 6th grade teachers respectively.) Group discussions were most cited as evidence at the 6th grade level, and writing or drawing samples were used by 50% of all teachers with little difference by grade level.

Student work samples were the most common attachments included by participating teachers with the Record of Reading Performance form. Fifty-four percent of all teachers included a sample of student writing or drawing as an attachment (61% of 3rd grade teachers, 42% of 4th grade teachers, 62% of 5th grade teachers and 45% of 6th grade teachers).

The types of student work samples included: book summaries and reports, close worksheets (the format used by the DRP test), writing samples in which the children described, commented or gave an opinion regarding their book selection, monthly book sharings, letters written by students to each other and to the teacher, children's journals and response notebooks which included student's own written responses to books as well as exercises such as sentence starters (e.g. If I were ... ; I never knew ... ; I wonder ... ). Students also recreated images they got from texts through drawing or painting or preparing posters about the book.

Formal records were made of oral interviews through a "Miscue Analysis" where teachers recorded and analyzed the mistakes a child made while reading aloud. A blank copy of a modified miscue analysis form developed by the Center for Educational Options (adapted from Ken and Yetta Goodman) is attached as Appendix M. These forms were included as attachments most often by teachers at the middle grade levels (48% of 3rd grade teachers, 42% of 4th grade teachers, 25% of 5th grade teachers and 9% of 6th grade teachers).

Teacher and facilitator notes supplemented the Record of Reading Performance and attachments with comments on the individual student conferences. Story re-telling and children reading aloud to teachers were the most common methods reported in teacher notes to determine comprehension. Conferences were also used to conduct miscue analyses, to evaluate strategies and skills, to elicit connections to personal life experiences, and to observe the child's behavior during oral reading. Finally, teachers found useful evidence when students participated in large and small group conversations, discussions and "book talks."

- "The individual conference allowed the teacher to ask for main ideas and check for vocabulary knowledge and difficulties. During this time vocabulary, comprehension, work habits and oral reading abilities were evaluated. Group sharing gave the teacher the sense and feeling the children had about the book and allowed the children to reveal their attitudes and interest." -- facilitator notes

Teacher notes and attachments also included information about how teachers knew their students could read the books they chose. In the following example a formal miscue analysis was included among the back-up materials for a third grader. This student read Just Listen, a story about a girl listening to city sounds. Teacher notes on the miscue form include: "He self-corrects a lot. He tried to read all the difficult words by reading just the first syllable. If a word is difficult for him to read, he just kept on reading. After encountering the same words many times in the story, he could read the word correctly. He was able to re-tell the story and while he was summarizing the events that happened, he told me about the things he heard at night when everyone in his house is already asleep." The miscue analysis gave the teacher the

opportunity to determine that the child could comprehend and reflect on the book, as well as to note the child's strategies for approaching the text.

Another teacher listed ways her students wrote about books. Students wrote to each other about what a book made them think about, how it made them feel, how they would do things differently than book characters or whether the book was worth reading. They also wrote questions to the author and compared characters in one book to another. Each written piece was evidence that the children comprehended what they read and were able to extend the reading experience in a personally meaningful way.

Many facilitator reports included reflections such as this one: "How do you know that the child comprehends a book? The child would talk about the book's topic, ask for other books on the same topic, and ask the teacher questions directly related to the book. The child will often offer information on the book; it might remind the child of something."

## What teachers learned about individual student strategies and attitudes

Teachers noted a wide variety of literacy strategies and skills used by students. These ranged from fairly specific "word attack" skills used by children to figure out how to read the words on the page to substantial interpretative strategies. This information enabled teachers to determine which skills students already used fluently and which should be the focus of further literacy instruction.

Strategies and skills observed included:

- recognizing words from beginning and ending sounds
- use of phonics to decode (or sound out) words
- guessing, making inferences or using picture clues
- use of context word clues to get meaning from print
- making word substitutions or self-corrections to maintain meaning while reading text aloud
- predicting outcomes
- writing detailed summaries of books or describing sequence of events, plot, characters
- character analysis, including discussion of interactions between characters (e.g. "She spoke of the struggle that went on throughout the book and the triumph at the end of the story" -- teacher quote)
- expressions of empathy for characters in story
- relating story to own personal experiences (e.g. "(Child) was able to relate story to herself, family, friends and heritage." -- teacher quote)
- discussion of social issues that occurred in the books

Teachers also noticed and reported that some children had little or no strategies when confronted with difficult words, or that their strategies were dysfunctional, e.g. being so involved in getting words "right" that they missed the point of the larger passage. Patterns were noted such as whether or not a student was easily discouraged by difficult reading material, and whether the child was afraid to make mistakes or took lots of risks when reading.

The individual conferences gave teachers an opportunity to piece together in a systematic way a fuller picture of their students as learners and whole persons. They enabled teachers to gather information about students' knowledge base, the reasons why children read (such as for pleasure or information), how children related books to their own personal experiences, interests and concerns, and issues related to reading at home. Teachers reported finding out that students enjoyed being read to, re-telling stories or sharing information. Many children also loved reading aloud to an adult or to the class or reported enjoying reading outside of school. Other teachers reported discoveries that helped explain a student's difficulties with reading, for example, one student whose father used reading assignments as punishments.

As a result of this close look at individual children, teachers observed students approaching reading in unique ways. One teacher noted four different approaches to texts among the students she studied; one student constantly questioned herself while she was reading, another self-corrected, another needed verbal discussion before answering written questions, and one had to read aloud. A facilitator described students reading for different purposes; some students loved to see images and picture everything in their minds, others were very interested in issues and making connections to their own lives, others were very aware of the emotions brought up by the books, and still others liked to get absorbed in a good plot.

Teachers who participated in the Multiple Forms of Evidence study consistently reported that they learned more about individual students by working one on one. As one teacher put it, "I found out things about children I didn't even know were there to find out." Many developed a better understanding of which aspects of a child's reading were weak and needed to be improved or found that they had underestimated the ability of some of their students. They took the opportunity to understand in a very concrete way how their students felt about reading. For instance, one student told a teacher, "I feel like I went in the book."

Student "subjects" felt good about being part of the study and getting the individual attention. For example, one teacher wrote that a special education student was "eager to participate" in the project. This individual attention itself seemed to lead to greater comprehension:

- "Children showed more interest in reading and comprehended more when involved in a one to one situation with teacher or peer." -- facilitator notes

## Implications for instructional practice

Teacher and facilitator notes reported specific instructional or assessment practices that they intended to emphasize as a result of their participation in this study. Examples of these comments include:

- "I've learned that when children are asked to think deeper by writing to each other or writing responses to the text they read better and enjoy reading more. They become more actively involved and are less passive readers." -- teacher quote

- "The teachers saw a tremendous advantage in miscue analysis as a way to analyze what 'strategies and skills' children have and then plan how to use the strategies instructionally. All the teachers felt it was a tool they would like to begin using at the beginning of the year to direct their instruction more precisely for particular children." -- facilitator's notes

Other instructional approaches mentioned by teachers as strategies to help children improve their reading were:

- more use of oral reading
- encouraging children to work in pairs
- direct teaching of self-correction strategies during reading
- asking direct comprehension questions
- re-telling and acting out stories
- test preparation aimed at helping children become as relaxed and comfortable in test taking situations as they were during classroom assessments
- use more literature in classroom instruction
- increased use of relevant and attractive non-fiction books

Through the individual student interview, teachers discovered which skills a particular student had absorbed and which needed additional emphasis in future instruction. The experience also helped teachers focus on skills to emphasize in the class as a whole. For example, one teacher noted that since mis-reading individual words did not necessarily interfere with student comprehension of a book, focussing on the overall meaning should be primary.

- "One teacher discovered that her guided reading strategies encouraged students to pay too much attention to details. They were having trouble getting the bigger picture." -- facilitator notes

The notes also convey a sense of how teachers planned to handle individual problems that might be uncovered through miscue analysis or any other strand of evidence. These included direct teaching of strategies, a greater emphasis on helping children go "deeper" into material

by talking or writing about the issues raised in the books, as well as the general statement "it was agreed that remediation should then take place."

Some teachers commented that they enjoyed working with students who were a challenge. Others felt inadequate to help non-readers. These teachers would be especially receptive to assistance in the use of particular strategies to build on the rich information about student strengths and weaknesses that is available from individualized assessments.

**Student choice of reading material**

One aspect of the study frequently emphasized by participants was the importance of student choice of reading material.

In most cases, children participated in selecting the books they read for this study. Students made their choices by interest, title, cover, illustrations, length of book or recommendation. Many students pursued sequels after enjoying a book. Teachers were interested to discover that an accurate picture of a child's reading ability was equally possible to obtain from a self-selected book as from a professionally-selected book or excerpt. In fact, when students were stimulated by self-selected material that interested them, they were even more likely to comprehend the material.

- "Children pick books that interest them which are more difficult than you would ever expect them to read. I still don't know how they get such good meaning." -- facilitator notes

- "Because the children selected their own book the literature pushes itself and all the students read willingly. In the past as the class read a book I always went over the questions provided at the end of each chapter. Next September I am not going to use those questions. Instead I am going to use the approach used in this project." -- teacher quoted in facilitator notes.

The significance of allowing children to choose their own book to read raised the pedagogical issue of developing children's range and repertoire of book choices. For instance, since books which had been read and liked by one student were recommended to others, some suggested that student sharing could be built into the instruction or that teachers could use their own love of books to steer children towards expanding their selection of reading material.

- " As children became familiar with the different writing styles of various authors, they developed preferences regarding authors. They grew more selective and when not satisfied with a particular book they returned it and requested another before proceeding any further. I like that." -- teacher quote

**Teacher comments about assessment**

Teachers appreciated the external recognition and encouragement provided by the study for collecting the types of evidence they, as professionals, find most useful and accurate.

- "When are they going to do this for all students? I have never taken test scores as the main basis for making any decisions about my students." -- teacher quote

- "How can we compare the answers a youngster fills in correctly in a dull text about camels (from the reading test) to what he/she knows or understands about love and loyalty after reading Stone Fox?" -- teacher notes

During the course of collecting evidence, teachers and facilitators involved in this study developed their own hypotheses and conjectures about the relationships between students, testing and reading. The patterns they noted turned out to be consistent across the entire sample studied.

- "Test scores aren't always accurate measures of achievement. For poor test takers, comprehension is often higher than teacher expectation." -- facilitator notes

- "Strong readers read books that are ranked below their mid-instructional DRP level and average readers tend to choose books closer to their mid-instructional DRP level." -- facilitator notes

**Collegial exchange**

Finally, participating teachers found that the opportunity to talk and collaborate with colleagues was a major benefit of the study. The regular meetings involving participating teachers and facilitators were occasions where observational records and samples of children's work were shared. Reviewing tangible data from the classroom gave substance to these discussions and fostered the development of a common language.

- "The most important aspect of this project was the opportunity for professionals to get together to discuss their students. My outside perspective sometimes challenged them and often reinforced ideas and concerns that they had felt but did not have the opportunity to voice." -- facilitator report

## Future plans

- "As with any positive 'pilot' experience, I and all three of my school teams view the Multiple Forms of Evidence as a practice session for adaptation and expansion next year." --facilitator report

Teams from many of the schools involved in the pilot plan to expand the ways they collaborate in the coming year. Participating in this study enabled groups of teachers to try out new assessment and instructional approaches, draw pedagogical conclusions and take risks. For example, one facilitator wrote that this project encouraged teachers to agree to heterogeneous grouping (i.e. grouping students of mixed achievement levels in the same classroom) for the first time next year.

During the November 1992 gathering of participants in the pilot study, teachers reflected on their experiences and indicated plans for the 1992-93 school year. Continuing Multiple Forms of Evidence project activities currently underway or being planned by school teams and by individual teachers include:

- Beginning project activities in the fall with administrative support, such as the assignment of common preps to participating teachers or scheduling regular team meetings, will enable teachers to observe childrens' literacy development throughout the year.

- Many teachers are interested in collecting data on increased numbers of students or even their whole class this year.

- Getting more teachers involved, either by teachers who were excited by the project bringing information directly to their colleagues, or through presentations at grade level and cross-grade conferences.

- Making use of readability values in selecting and organizing books for classroom use through projects such as developing suggested reading lists of books with readability values which match students' test scores or organizing classroom library collections by readability level.

- Making a formal comparison of childrens' reading performance of non-fiction vs. fiction by collecting samples of each type of material for each child being studied.

## Summary

Experienced as well as new teachers took advantage of the opportunity to take part in the Multiple Forms of Evidence Study to investigate reading and assessment in relation to individual

children. This study's serious review of classroom-based student performance data gave recognition to teacher knowledge and practices.

- "The teachers who participated all indicated that they have altered their views of themselves as teachers, and assessors, and of assessment in general. Much professional growth was achieved in a very short period of time, in both theoretical and philosophical underpinnings to professional practice as well as in the practice itself." -- facilitator notes

The information gained from collecting multiple forms of evidence of a student's reading ability is extremely valuable and immediately useful. The assessment practices used by teachers in this study not only helped teachers evaluate how well a child could read with understanding, but also gave them additional insights into the child's thinking and strategies, clarified their picture of specific strengths and weaknesses, and could lead directly to the development of instructional plans which responded to the observations.

Teachers who work with children every day are in the best position to evaluate their students' progress. The assessment strategies used in this study grew out of classroom practices and research and have the potential to be realistic options in most classrooms. The wealth of information that becomes available to teachers who use them is far too valuable to overlook. As one fifth grade bilingual education teacher observed, "With traditional tests, you teach the lesson not the student. But with this project, I get to know the students and whether I am getting through to them."

## Chapter Four

## Policy Implications

If we want accurate and complete assessments of students and schools, we must abandon our sole or even primary reliance on a single standardized test and adopt fuller, performance based assessment practices.

Replacing one test with another is not the answer. A reading-like activity in an artificial testing situation does not yield the same kind of rich and useful information that evaluating a child reading an actual book in a normal everyday setting provides. The 61 teachers in this pilot were clear on this point. No standardized test given once per year will yield accurate and complete assessment information.

Even the best performance-based assessment is inadequate if it is the only assessment strategy or if it is only employed once a year. Our assessment strategy must shift from one-shot instrumentation to approaches that systematically tap the everyday evidence of learning in the classroom over time and across settings. Such assessment approaches are curriculum imbedded and necessarily call for a much more significant role on the teachers' part.

At every level of governance, policy makers should insist that schools and districts not only collect multiple forms of evidence of student progress but also use this information to better serve their students. Schools and districts should be held accountable for collecting and using multiple forms of evidence of student progress to:

- Inform instructional strategies within the classroom;
- Travel with students to inform instructional and placement decisions as they move to a new grade or new school;
- Serve as one indicator of overall school effectiveness.

Just as no school should be deemed to be failing on the sole basis of test scores, no school should be assumed to be effective because of high test scores. One important measure of school effectiveness should be how well the school collects and uses classroom assessment evidence both to evaluate its overall performance and to respond to individual students' needs. Policy makers with responsibility for monitoring school quality should routinely look for such practices. Where low test scores and other indicators suggest that schools may not be doing enough to help students, observational and performance-based assessment data, pulled at random

from student files and studied by district and school system officials, should serve as a valuable double check on the teaching and learning actually occurring in the school.

A standardized test should never be the sole basis of a judgement about either an individual student or a school. Such a judgment hurts rather than helps accountability because a single test score provides information that is neither accurate nor complete enough to stand alone. This means that certain current practices must change:

- The decentralization law should be amended so that the New York City schools are no longer ranked each year based on their reading test scores;
- The state should revise its methods for identifying schools in need of improvement so that the performance on one test by one grade does not automatically result in that school being cited for poor performance;
- Schools and districts should no longer be permitted to use test scores as the sole basis for tracking students into classes grouped by achievement level or requiring students to repeat a grade.

Is it realistic to expect all public school teachers in New York City to systematically collect and record multiple forms of evidence of student progress? Based on the experience of the teachers who participated in this project, the answer is a qualified yes. Although the 61 teachers were all volunteers, many had little if any prior experience with "alternative assessment." With relatively little outside assistance, they were able to collect, analyze and use a variety of evidence of student progress. Indeed, five teachers were able to collect evidence for their entire class. With advice from these teachers, we have identified key areas where teachers and schools will need support if this relatively simple, common sense approach to assessment is to become a professional norm and systemwide practice.

## Professional Development

Teachers have three distinct professional development needs in regard to improving assessment:

1. Learning about the assessment methods themselves;
2. Learning to use the resulting information about students to improve instruction;
3. Working on classroom organization and management strategies to support collection of assessment data and to integrate assessment with regular instructional activities.

In our pilot, most of this work occurred within the school, in monthly meetings convened and facilitated by a trained OREA assistor or an SBM/SDM facilitator. District or school-based staff developers, principals or assistant principals can also play this convening role. These meetings offered invaluable opportunities for reflection and mutual support. They should be supplemented by opportunities for teachers to study with assessment experts, learn from more experienced colleagues and compare notes with other schools, as was provided in our project orientation and our follow-up meeting with participating teachers to review the draft of this report.

## Administrative support

District offices and building administrators played a critical role in supporting teachers' participation in this project. For instance, they rescheduled teachers' programs to create common prep periods so that teachers could meet together for staff development and reflection during the regular school day without detriment to their students. They provided substitute teachers so that teachers in the study could attend the half day orientation or spend a few hours working with individual students. They permitted participating teachers to use previously scheduled staff development days for assessment work, rather than other planned workshops.

To obtain this kind of support systemwide will require some consciousness raising and professional development for district and school administrators as well as teachers on the importance of sound assessment practices and in ways to reorganize existing resources to support them.

## Time

Even with the most enthusiastic support from principals and superintendents, finding time during the busy school day for staff development, planning meetings and experimentation with new teaching practices is the greatest challenge to any school reform effort. Many of the teachers in our study used their prep and lunch periods to assess individual students. Some of their colleagues, who expressed interest in this project, could not join it because they were already donating their preps and lunches to other school improvement efforts.

However, teachers who managed to collect classroom evidence on their entire class provided some important insights on how to find time. They observed that when they were first mastering the assessment techniques, they needed to start with just a few of their students and appreciated an occasional hour or two of coverage so that they could work with just one child and leave the rest of their class in another teacher's care. However, once they were comfortable with the assessment methods, they found it was possible to work with one child during a scheduled classroom silent reading period or to concentrate on taking notes on one child's response during a group discussion of a book. One teacher reported that using these simple strategies, she was able to get reading samples on her entire class during a three week period.

The key, she noted, was that during those three weeks, she made individual assessment interviews her top priority.

Other teachers suggested that out-of-classroom support persons such as Chapter One reading teachers, resource room teachers, school-based staff developers, assistant principals and cluster teachers could all play a valuable role both in helping collect classroom assessment and working with a class while the classroom teacher assessed individual students.

The Fund gave one school a $500 grant to pay for substitute teachers so that three teachers, who had not been active in previous alternative assessment pilots, could study their entire classes. A fourth teacher, who had already used these techniques, volunteered to study her class without any special assistance, as did an experienced teacher at another school. Thus, this modest investment gave teachers the time to try out techniques that once mastered, they can employ without any special supports.

Perhaps the greater challenge is to find time for teachers to meet together for professional development and mutual support. In our study, these meetings also tended to take place during prep periods and lunches which limited participants to those who happened to share common free periods. Participants had some additional suggestions on how such meetings could be facilitated:

- Districts should provide meeting time for teachers engaged in similar classroom assessment efforts during regularly scheduled half-day staff development programs.

- Schools should encourage teachers to use part of the three days of professional development time provided in their contract to participate in assessment workshops such as those offered by the Accountability Project and to visit other classrooms and other schools where similar classroom assessment efforts are in progress.

- A series of meetings such as those led by OREA assistors and SBM/SDM facilitators in our project could be incorporated into the in-service credit conferring program for newer teachers offered through the Division of Instruction and Professional Development.

## Policy priorities regarding testing

The past two decades have seen significant advances in our understanding of children's literacy development. If we expect teachers to change their classroom practices by building upon this new knowledge, it is incumbent upon districts and policy makers also to bring assessment practices into line.

In the view of participating teachers, the single greatest obstacle to sound assessment and instruction practices is the intense pressure on students, teachers, principals and superintendents to get high scores on the annual standardized citywide reading test. Every year, schools are ranked from best to worst based solely on how many of their students perform "at or above grade level" on the reading test. (This ranking is actually required by an obscure provision in the state decentralization law regarding teacher assignments.) Test results in certain grades are a key factor in State Education Department decisions to place certain schools "under review." Everything from principal tenure to real estate values around a neighborhood school can ride on the results of one test.

A very concrete example of the effect of over-reliance on standardized test scores was called to our attention by several teachers who reported that in their districts, policies required teachers to give students poor reading grades, regardless of their actual reading performance, if they score low on the once-a-year, multiple choice test. Naturally, these teachers felt the policy was unfair and hurtful to those students who read better than their test scores would predict. Such policies are, in part, a response to state education regulations that require parents to be notified and a remediation plan to be in place if students score below a certain point on the third, sixth or eighth grade administrations of the DRP. Teachers may be required to base grades solely on test scores to avoid a situation where parents get a good report card from the teacher and then receive a letter from the principal warning that their child needs help in reading.

We know that judging schools on the basis of a single test actually reduces accountability and is detrimental to instruction. Too much emphasis on one test forces schools to teach to that test, neglecting other important subjects, concepts and skills and placing undue emphasis on what can be tested in a once-a-year multiple choice format. We know from this pilot and from the warnings on every test manual, that individual students should also not be judged on the basis of one test score. The Chancellor has come out strongly against the ranking of schools by a single test score. His staff, particularly at the Office of Research, Evaluation and Development, has played an instrumental role in this study and in other efforts to encourage sound student assessment practices. The school system is piloting new early childhood reading assessments and middle school math assessments citywide. Over 50 schools are engaged in piloting the Primary Language Record, a comprehensive assessment system for elementary school students in reading, writing and oral communication developed in London.

The Commissioner of the State Education Department, in The New Compact for Learning, has also demonstrated his support for sound, performance-based assessment of student progress. At the national level, the New Standards Project and the National Assessment of Educational Progress, among other efforts, are bringing the best minds together to pilot better assessment practices. Across the country and here in New York City, an enormous amount of work has been done to develop rich and comprehensive assessment methods and use them in our classrooms.

Yet every year students and schools continue to be judged on the sole basis of test scores. How do we get out of this quandary?

In addition to the overall policy recommendations made above, this study has yielded some simple, practical and cost effective ways to restore standardized tests to their rightful place as one of several indicators of student progress.

- Redesign report cards to encourage teachers to report their classroom assessment findings, and where necessary,compare test scores to classroom evidence. There is no need to record DRP readability values for all children. But when the teacher suspects that a test score is under-reporting the child's actual reading level, looking up and reporting the DRP readability values of books the child can actually read is a helpful way to translate classroom evidence into the language of test scores and draw honest and responsible conclusions.

- Report test information in "Degrees of Reading Power" units, with examples of passages that had readability values equivalent to the test scores, rather than only in percentiles which provide no concrete information to either parents or teachers.

- Reduce the time spent on test practice. Many of the teachers in this study report that they are now asked to spend the equivalent of weeks of classroom time on the administration of practice tests. In some districts, practice for the test begins a month before the actual test. In others it is required beginning in September. In other words, during time that could be spent reading, discussing and writing about real books, students are instead required to answer multiple choice questions about passages taken out of context, in order to maximize their spring test scores. The teachers in our study believe this time would be better spent in instituting a well-rounded classroom assessment program, integrated with instruction.

## Conclusion

Using multiple forms of evidence is more complicated and less convenient than relying on a single test score. But it is also a more accurate and therefore, more responsible approach to assessment. After all, our primary goal is not to produce students who do well on standardized, multiple choice tests. Our goal is to produce students who can read with enjoyment and understanding and who can talk and write thoughtfully and clearly about what they have read. Common sense tells us that the best way to know whether our students can read is to listen to them read, talk with them about what they have read and read what they write about what they have read. No single form of assessment, however convenient, can replace evaluation of student work in different contexts, over time. In the interest of our children, let common sense prevail.

# APPENDIX A

**ASSESSMENT OF CHILDREN'S LITERACY: MULTIPLE FORMS OF EVIDENCE PROJECT**
**1992-93**
**Letter of Commitment**

Please complete this form and return it by February 28, 1993 to Maureen Houtrides, Office of Research, Evaluation, and Assessment, 110 Livingston Street, Room 735, Brooklyn, N.Y. 11201. The number of schools that will be accepted as participants in the project is limited so it is important that you return this form as soon as possible. If you need any further information, please call Maureen Houtrides at (718) 935-3284.

1. School:_____ 2. District:_____

3. Address:_____

4. Contact Person:_____ 5. Telephone:_____

6. Name of your Accountability Project assistor and/or SBM/SDM facilitor:

_____

7. Indicate the initiative(s) your school is participating in:

_____Chapter I Schoolwide _____Corridor

_____Other SBM/SDM _____Other:_____
(Specify)

Letter of Commitment

Our SBM/SDM team has discussed the Assessment of Children's Literacy: Multiple Forms of Evidence Project and agrees that:

- at least three classroom teachers in grades three through six will participate from our school;
- participating teachers will each collect information on three or four students;
- participating teachers will meet at least once a month with their assistor or facilitator to present evidence they have gathered on individual children and to discuss what they have learned; and
- the administration will provide a common prep or meeting time for participating teachers.

## APPENDIX A
continued

We agree that our school will share its systematic collection of reading performance evidence with our Accountability Project assistor and/or SBM/SDM facilitator and other participating schools so that everyone's knowledge from this pilot can be strengthened.

Assessment of Children's Literacy: Multiple Forms of Evidence Project

PARTICIPANTS

| Name (Printed) | Signature | Grade |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

Principal_____

SBM/SDM Chair_____

UFT Rep._____

Parent Rep._____

THANK YOU FOR YOUR PARTICIPATION

## APPENDIX B

### Participants in Multiple Forms of Evidence Study

Small Group Study

| Grade | # of Teachers | # of Students |
|---|---|---|
| 2 | 2 | 2 |
| 3 | 23 | 53 |
| 4 | 12 | 31 |
| 5 | 8 | 22 |
| 6 | 11 | 31 |
| Subtotal small groups (16 schools) | 56 | 149 |
| **Whole Class Study** | | |
| 3/4 | 1 | 25 |
| 5/6 | 4 | 94 |
| Subtotal whole class (2 schools) | 5 | 119 |
| **Total Participants:** 18 schools | 61 teachers | 268 students |

# APPENDIX C

## Record of Reading Performance

| TITLES OF BOOKS | EVALUATION: READS WITH EASE & UNDERSTANDING | EVALUATION: READS WITH ASSISTANCE | DRP READABILITY | NATURE OF EVIDENCE: * ORAL READING | INDIVIDUAL CONFERENCE | GROUP DISCUSSION | WRITING/ DRAWING SAMPLES | OTHER | COMMENTS * |
|---|---|---|---|---|---|---|---|---|---|
| Ramona Quimby, Age 8 | ✓ | | 54 | ✓ | | | | | See attached miscues 5\|5 + 5\|14 |
| Miss Nelson is Missing | ✓ | | 45 | | ✓ | | | | See notes attached |
| Ramona the Brave | ✓ | | 52 | | ✓ | | ✓ | | See attached |
| Curious George gets a Medal | ✓ | | 47 | | ✓ | | | | See attached notes |
| | | | | | | | ✓ | | 2 Writing/Drawing samples attached |
| | | | | | | | | | |
| | | | | | | | | | |

38

District 7 School P.S. 30 Child's Code _____

Child's Grade 4 Gender: F ✓ M ____

Is English the child's only language ? Yes ✓ No ___ Don't Know ___

* Attach your notes and student work samples, if available.

Student's DRP scores for Spring 1992:

41

| RS | IND | LW | MID | HI | FRS | NP | NOS | ST |
|---|---|---|---|---|---|---|---|---|
| 46 | 38 | 46 | 49 | 52 | 60 | 54 | 52 | 5 |

45 V3

1991

32 MID

24 NP

30F4F2

P.S.. 30x
Class 4-2

May 28, 1992

Tittle Ramona the Brave
Illustrated by Alan Tiegroen
Author- Beverly Clery

This Chapter is about Ramona's Sister. She was pushing Ramona on the swing and someboys started calling Ramona's Sister Jesus Beezus and Beezus Jesus. A Romena got out the swing and started t preach. She embarrassed her sister. And they marched home and told frier mother. Beezus Said She did not wat to be called Beezus again. So they tryed To give her a new name.

**APPENDIX C**
continued

MISCUE ANALYSIS FORM

Student _____ School P.S 30

Date of Birth 10/22/82 Teacher T. Caraballo

Grade 4th Date of Reading 5/5/92 +5/11/92 Level DRP Unit 54

Book Read Ramona Quimby, Age 8

Author Beverly Cleary Was book familiar to child? No

Who chose the book? Child ✓ Teacher ___

During Reading:

Interest (Was the child relaxed or anxious? Interested? To what extent is this sample representative of the child's skill and attitude regarding reading?): She was very nervous. She didn't like being taped. She feels she does much better when she is nervous. This is a good sample of her skill.

Understanding (e.g., intonation, spontaneous remarks, use of picture clues, self-correcting for meaning, responses to questions):
Self-corrects a lot. Reads with intonation

Strategies (e.g., willing to guess, maintains momentum, seeks support from teacher, monitors for accuracy): She maintains momentum, willing to guess. Will reread to try to better understand what she is reading.

During Retelling:

Capture the sense of the story? (plot, action, conclusion):
Was able to retell the story. She didn't like chapter 1 very much. She liked reading chapter 6 Extremely detailed about the throwing up incident.

Discuss the characters? (interactions, changes):
She is critical of Ramona. feels she is too nervous. Feels she is too hard on herself. She had a good understanding of the characters.

Connect the story to her/his own experiences?:
"You shouldn't be called a nuisance for the things Ramona did" "You shouldn't feel embarrassed if the same thing happens to you."
She ~~said~~ has never thrown up in school, but feels if it ever happened she wouldn't be embarrassed

**APPENDIX C**
continued

30F4F1

During Retelling (continued):
Approach text critically? (offer opinions, raise questions, etc.):

1) When you are reading and you come to something you don't know, what do you do? Do you ever do anything else?
I try to sound out the word. I read the whole sentence then go back using clues in the sentence to try to read the word.
2) How did you learn to read?
My kindergarden teacher taught me to read.

3) What kind of a reader are you?
a normal reader. "I read like a fourth grader."
Other Remarks:

Instructional Implications:

Calculations:
Accuracy rate:

$$\frac{\text{total words - total miscues + self-corrections}}{\text{total words}} \quad _____\%$$

Meaningful miscue rate: $\frac{\text{meaningful miscues}}{\text{total miscues}}$ _____%

Self-correction rate: $\frac{\text{self-corrections}}{\text{total miscues}}$ _____%

APPENDIX D

## Table 2. Readability of Prose Samples in DRP Units

### 34 DRP Units

Bears are big. They need a lot of food. Bears eat meat. They eat bugs. They eat berries. They eat honey. They eat fish, too. Bears feed in the spring. They feed in the summer. They feed in the fall. Bears look for food then. They hunt. They fish. They dig roots. They pick berries. They eat a lot. They grow fat. Soon, winter comes. It gets cold. It snows. But the bears don't need to go out. They don't need food. They are fat enough. They can sleep.

### 39 DRP Units

A bird's wings are well-shaped for flight. The wing is curved. It cuts the air. This helps lift the bird. The feathers are light. But they are strong. They help make birds the best fliers. A bird can move them in many directions. Birds move their wings forward and down. Then they move them up and back. This is how they fly.

### 43 DRP Units

Many states are dry in summer. They get hardly any rain. Nearly all their water comes from melted snow. It is stored. It is kept in dammed-up ponds and lakes. It is used during the growing season to water farms and orchards. Farmers buy the water. They are told how much they will be able to get. The amount changes each year. It depends on how snowy the winter was. A farmer needs to know how much he will receive. It allows him to decide which of several crops he ought to plant. The choice is based on how much water different crops need.

### 47 DRP Units

The part of a beach between high and low tide is called the middle beach. It is home to many plants and animals. But life on this middle beach is hard. There is no protection against the wash of the oncoming waves. Some animals survive by digging holes in the sand. They can stay in their homes under ground. The undertow will not pull them out to sea. They are safe.

### 51 DRP Units

Most creatures take care to protect their eggs. The walking stick does not. It just drops its eggs, scattering them loosely on the ground. Dozens and dozens drop at a time. As the eggs fall onto dry leaves, they sound like raindrops falling. Many of the eggs do not hatch. But enough do so that the walking sticks will not die out. They have existed on earth since before the era of the dinosaurs.

### 56 DRP Units

The people of Greece used the alphabet of the Semites. At first the Greeks wrote from right to left and left to right in alternating lines. The Greek name for this system of writing came from their words for "ox" and "turn." The method reminded them of oxen going back and forth, plowing a field. Eventually, the Greeks wrote only in one direction, as most people do now.

### 60 DRP Units

The ouija board is a simple rectangular piece of wood. All the letters of the alphabet are set out in a semicircle across a long edge. The ten digits and the words "yes," "no," and "goodbye" appear below. A small heart-shaped piece of wood called a planchette is mounted on casters so it can move easily on the board. When one places his fingertips lightly on the planchette, it slides around. It moves apparently without any conscious control on the part of the operator. Its pointer is supposed to spell out the answers to questions.

### 64 DRP Units

Wall paintings are especially vulnerable to atmospheric change. Archaeologists know this. Hence they try to discover, before opening a tomb, whether they will find murals. Special tools have been designed for this purpose. One of the most useful is a kind of camera that can be dropped into the ground before the digging starts. If the camera indicates the presence of wall art, scientists can prepare to take steps to preserve the painting as soon as it is reached.

### 73 DRP Units

Hellenistic literature showed an interest in individual history and psychology, rather than man in general. Theophrastus' *Characters*, with its detailed portraits of such types as the flatterer, appeared during this time. Biography, dealing with the lives of real people, was a flourishing form. And in philosophy the emphasis was on personal conduct rather than speculation about reality.

### 81 DRP Units

Jefferson's preference for an agrarian society and his idealization of the independent farmer reflected a conviction that representative government required a secure and relatively prosperous economic base to function successfully. He perceived the farmer as economically independent, and thus unlikely to surrender his judgment as a citizen to the influence of demagogues. His dislike and distrust of cities derived from a conviction that urban conditions, especially for the poorer classes, forced men into such a bitter struggle for sheer self-preservation that their natural moral sense could not be relied upon to produce social harmony or to guarantee responsible citizenship.

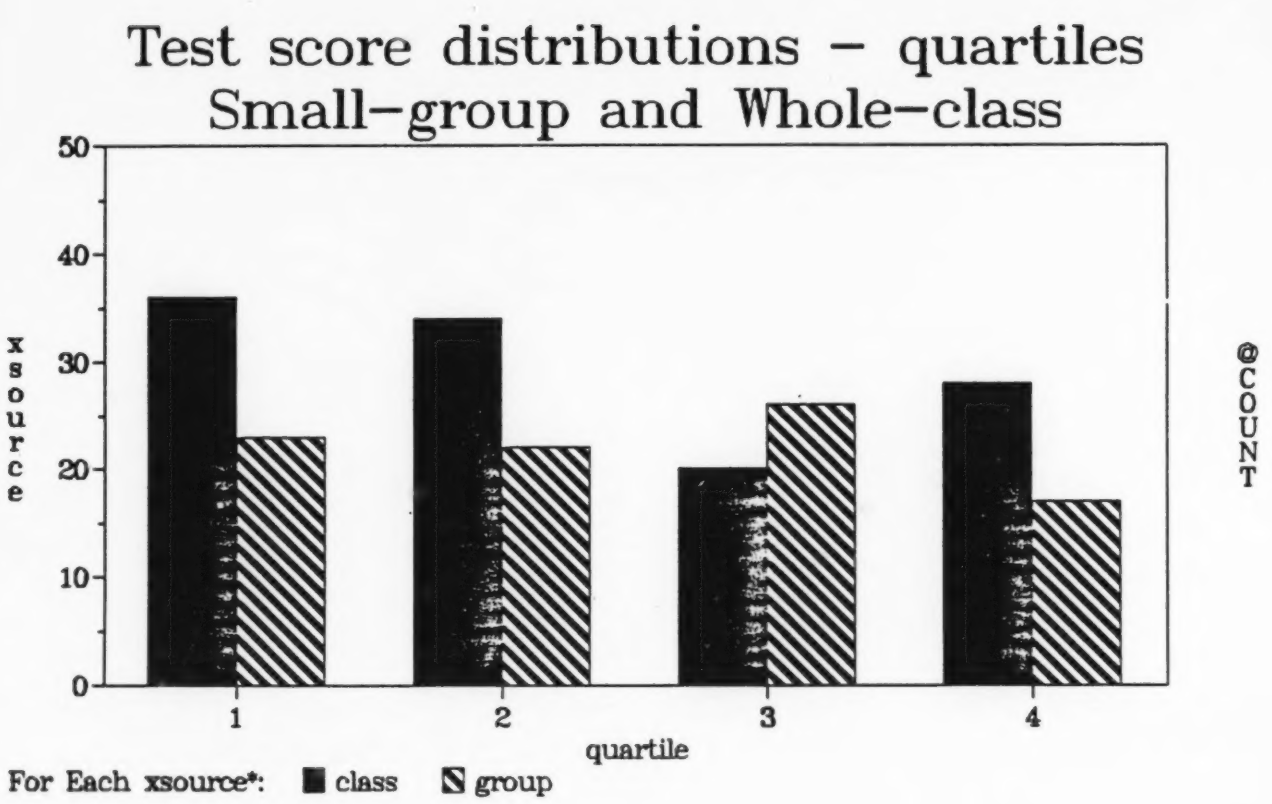
---

Note: The readability calculations are based upon longer samples.



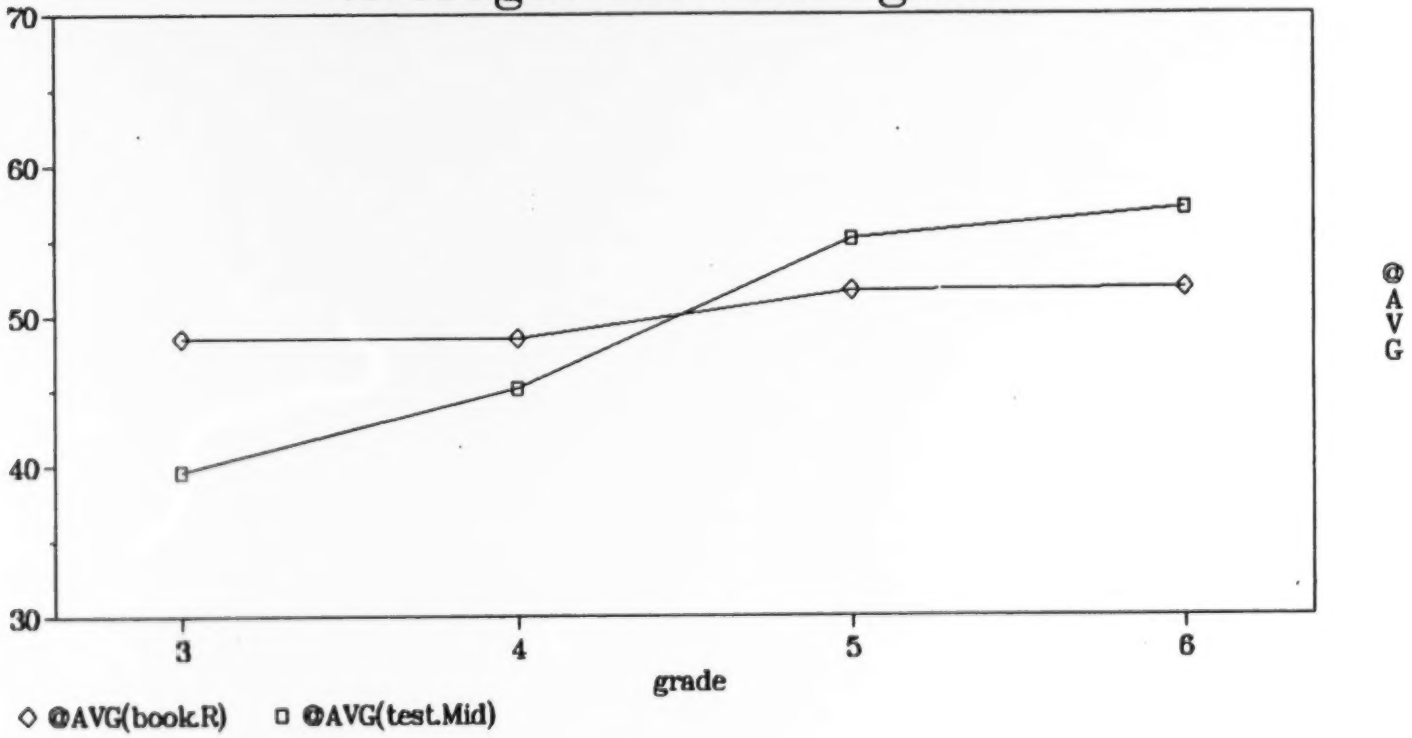
BEST COPY AVAILABLE

**APPENDIX E**


Test score distributions – quartiles
Small–group and Whole–class
50
40
30
20
10
0
xsource
1
2
3
4
quartile
@COUNT
For Each xsource*: class group

**APPENDIX F**

# Readability & Test Scores [in DRP units]
## Averages for each grade



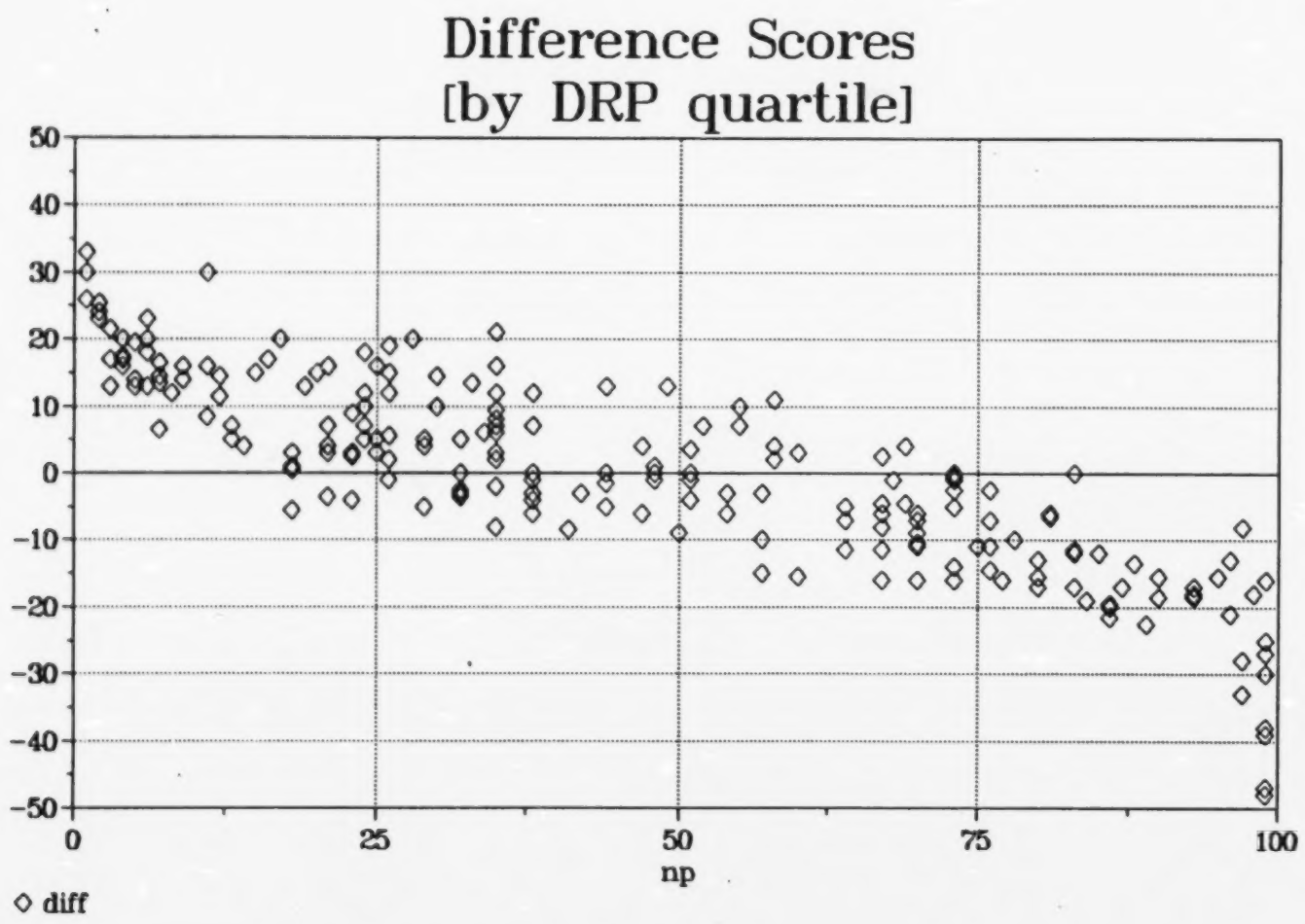
# APPENDIX G

**Difference Scores:**

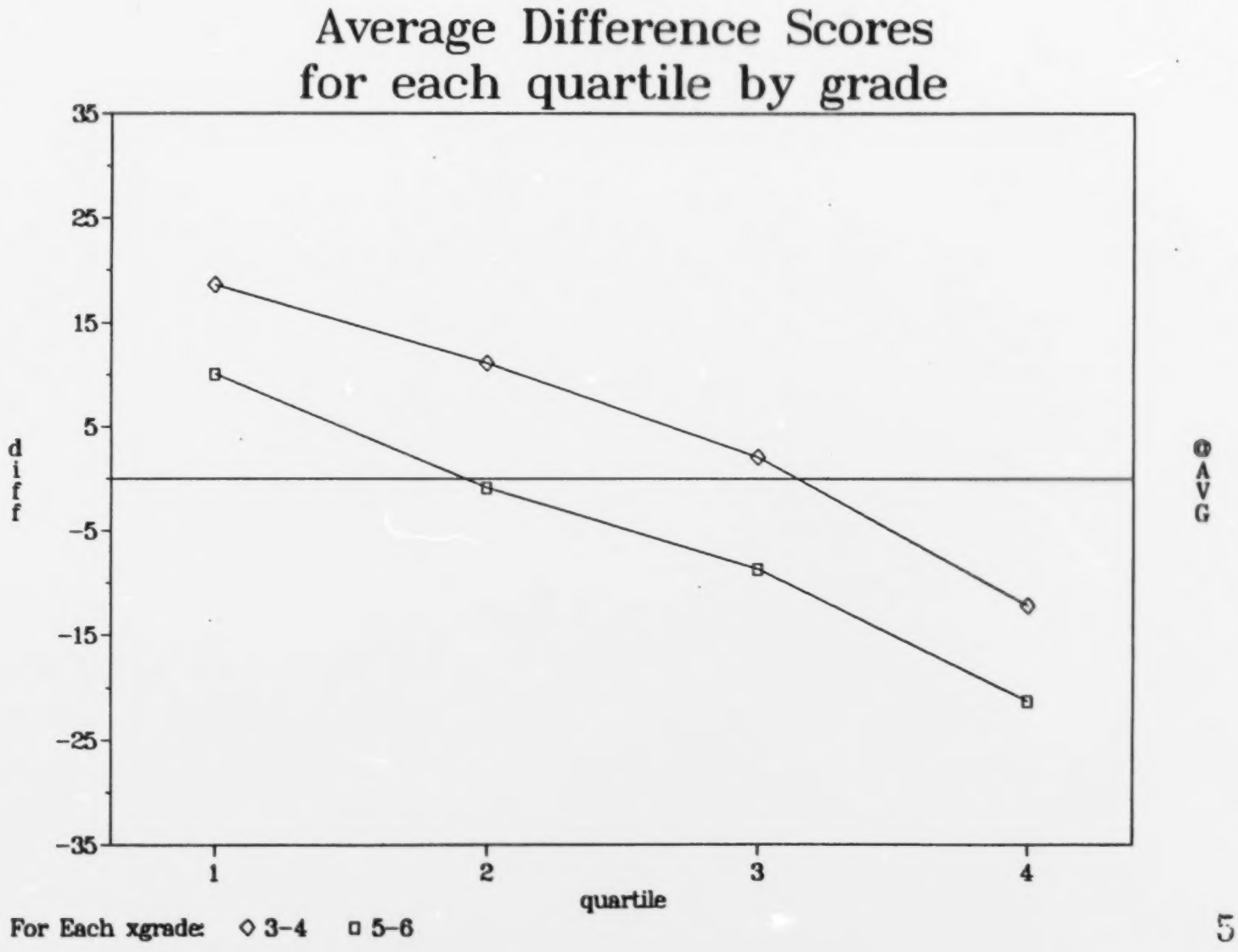
Difference between the DRP readability value of books and the students' test scores reported in DRP units (mid-instructional score)

| Difference in DRP units | Grades | | |
|---|---|---|---|
| Readability - Test | 3-4 | 5-6 | ALL |
| | % of students | % of students | % of students |
| 20 or higher | 14.10 | 3.10 | 7.25 |
| 11 to 19 | 30.77 | 11.63 | 18.84 |
| -10 to 10 | 44.87 | 55.04 | 51.21 |
| -11 to -19 | 7.69 | 21.71 | 16.43 |
| -20 or lower | 2.56 | 8.53 | 6.28 |
| TOTAL | 100.00 | 100.00 | 100.00 |

**APPENDIX H**


Difference Scores
[by DRP quartile]
50
40
30
20
10
0
-10
-20
-30
-40
-50
0
25
50
75
100
np
◇ diff

**APPENDIX I**


Average Difference Scores
for each quartile by grade
35
25
15
5
-5
-15
-25
-35
diff
1
2
3
4
quartile
For Each xgrade: ◇ 3-4 ▫ 5-6
AVG

# APPENDIX J

## Comparison of Aggregate Data
## Third Grade Sample

DRP Norm-Referenced Test Scores

| | Below Average * | At or Above Average | Total |
|---|---|---|---|
| # of children | 32 | 20 | 52 |
| % of children | 62% | 38% | 100% |

* Average is defined as the 50% percentile.

DRP Readability values of books students could read compared to average DRP level of state sample

| | Below Average * | At or Above Average | Total |
|---|---|---|---|
| # of children | 21 | 31 | 52 |
| % of children | 40% | 60% | 100% |

* Average readability values defined as 48 DRP units, based on averages reported for a state sample of third-grade instructional materials (DRP guide).

# APPENDIX K

## MULTIPLE FORMS OF EVIDENCE

SMALL GROUP STUDY
NATURE OF EVIDENCE
Reported on the Record of Reading Performance

| Grade (#) | Oral Reading | | Individual Conference | | Group Discussions | | Sample | | Other | |
|---|---|---|---|---|---|---|---|---|---|---|
| | # | % | # | % | # | % | # | % | # | % |
| 3 (23) | 22 | 96% | 19 | 83% | 5 | 22% | 11 | 4% | 8 | 35% |
| 4 (12) | 12 | 100% | 11 | 92% | 7 | 58% | 6 | 60% | 0 | 0% |
| 5 (08) | 7 | 87% | 8 | 100% | 3 | 37% | 5 | 62% | 2 | 25% |
| 6 (11) | 9 | 82% | 9 | 82% | 7 | 64% | 5 | 45% | 2 | 18% |
| Total (54) | 50 | 92% | 47 | 87% | 22 | 41% | 27 | 50% | 12 | 22% |

## APPENDIX L

**Small Group Study**
**Teacher Attachments**

| Grade (#) | Book Excerpts | | Miscue Analysis | | Sample | | Interview | |
|---|---|---|---|---|---|---|---|---|
| | # | % | # | % | # | % | # | % |
| 3 (23) | 9 | 39% | 11 | 48% | 14 | 61% | 6 | 26% |
| 4 (12) | 4 | 33% | 5 | 42% | 5 | 42% | 3 | 25% |
| 5 (8) | 2 | 25% | 2 | 25% | 5 | 62% | 3 | 37% |
| 6 (11) | 2 | 18% | 1 | 9% | 5 | 45% | 1 | 9% |
| Total (54) | 17 | 31% | 19 | 35% | 29 | 54% | 13 | 29% |

# APPENDIX M

## MISCUE ANALYSIS FORM

Student_____Date of Reading_____

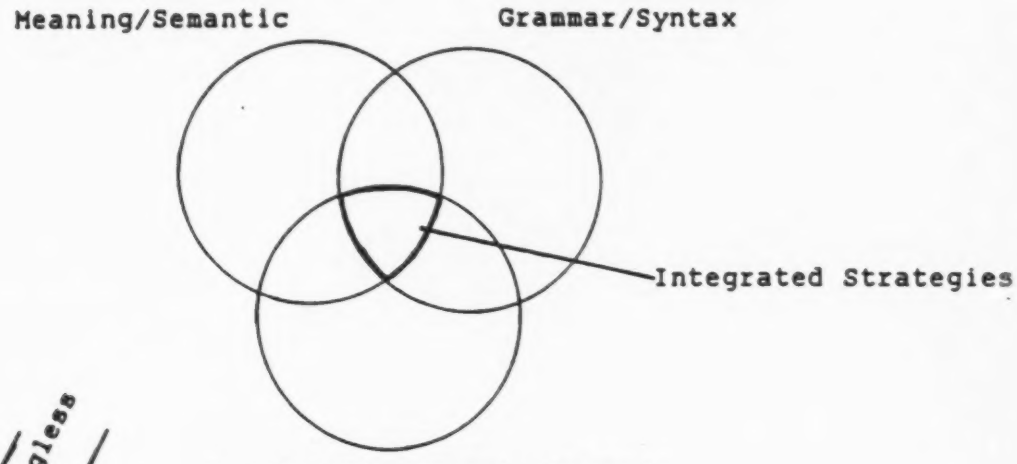
Date of Birth_____Grade_____School_____

Teacher_____Who chose the book? Child Teacher

Book Read_____

Author_____Was the book familiar to child?_____

### Cueing Systems and Reading Strategies




| self-correction | meaningful miscue | meaningless miscue | Reader | Text |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

**During Reading:**
**relaxed or anxious?:**

**interested?:**

**intonation?:**

**spontaneous remarks?:**

**use of picture clues?:**

**self-correcting for meaning?:**

**willing to guess?:**

**maintains momentum?:**

**seeks support from teacher?:**

**monitors for accuracy?:**

**During Retelling:**
**Capture sense of characters and story?:**

**Child Interview:**
1. What kind of reader are you?

2. How did you learn to read?

3. When you are reading and you come to something you don't know, what do you do? Do you ever do anything else?

**Instructional Implications:**

**Calculations:**
**Meaningful Miscue Rate:** $\frac{\text{meaningful miscues}}{\text{total miscues}}$ _____%

**Self-correction Rate:** $\frac{\text{self-corrections}}{\text{total miscues}}$ _____%

# APPENDIX N

## Record of Reading Performance

| TITLES OF BOOKS | EVALUATION | | DRP READABILITY | NATURE OF EVIDENCE * | | | | | COMMENTS |
|---|---|---|---|---|---|---|---|---|---|
| | READS WITH EASE & UNDERSTANDING | READS WITH ASSISTANCE | | ORAL READING | INDIVIDUAL CONFERENCE | GROUP DISCUSSION | WRITING/ DRAWING SAMPLES | OTHER | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

* Attach your notes and student work samples, if available.

District _____ School _____ Child's Code _____

Child's Grade _____ Gender: F _____ M _____

Is English the child's only language ? Yes ___ No ___ Don't Know ___

Student's DRP scores for Spring 1992:

| RS | IND | LW | MID | HI | FRS | NP | NCE | ST |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |